Collaboration to Clarify the Costs of Curation

ADA 2014
WELCOME TO SUMMER SCHOOL ON ACCESS TO DIGITAL ARCHIVES
4th July 2014, Split, Croatia

The Costs of Curation – are we on the right track?

Neil Grindley
n.grindley@jisc.ac.uk
Head of Resource Discovery

Jisc Coordinator,
The 4C Project
Friday, 4 July 2014

<table>
<thead>
<tr>
<th>TIME</th>
<th>LECTURER</th>
<th>TOPIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.30-11.30</td>
<td>Neil Grindley</td>
<td>The Costs of Curation – Are we on the right track? - lecture</td>
</tr>
<tr>
<td>11.30-12.00</td>
<td></td>
<td>Coffee break</td>
</tr>
<tr>
<td>12.00-13.00</td>
<td>Neil Grindley</td>
<td>The Costs of Curation – Are we on the right track? - group work</td>
</tr>
<tr>
<td>13.00-14.00</td>
<td>Neil Grindley</td>
<td>The Costs of Curation – Are we on the right track? - discussion</td>
</tr>
<tr>
<td>14.00-15.30</td>
<td></td>
<td>Lunch break (on your own)</td>
</tr>
<tr>
<td>15.30-17.00</td>
<td></td>
<td>Round table - participants discuss their future research</td>
</tr>
<tr>
<td>17.00-17.15</td>
<td></td>
<td>Coffee break</td>
</tr>
<tr>
<td>17.15-18.45</td>
<td></td>
<td>Closing session and announcement of ADA 2015</td>
</tr>
</tbody>
</table>
Project Summary

The Collaboration to Clarify the Costs of Curation (4C) project will help organisations across Europe (and beyond) to more effectively invest in digital curation and preservation.

Vision

The 4C vision is to create a better understanding of digital curation costs through collaboration.

Mission

Our mission is to provide useful, useable resources which support the process of cost management in digital curation.
Assessment

*Tasks*
- Assess cost models & strategies
- Examine good practice
- Analyse requirements
- Integrate components
- Produce guidance & briefing materials
- Setup costs exchange

Enhancement

*Tasks*
Examine and refine related concepts
- Value
- Risk
- Benefits
- Sustainability
- Economic Reference Model

Engagement

*Tasks*
- Engage stakeholders
- Raise awareness
- Organise meetings
- Promote Research & Innovation
- Build community network

Networking & Coordination

Affiliate Partners & Stakeholders

Project Coordination

*Tasks*
- Project meetings
- Project reporting
- EC liaison
- Budget oversight
- Outputs QA

Outputs

Reports for General Dissemination

Curation Costs Exchange

Events, Workshops, Meetings & Reports

Collaboration to Clarify the Costs of Curation

Jisc

Reports for European Commission

Submission of Roadmap to the EC

Curation Engagement Tasks

Engagement Tasks

Assessment Tasks

Enhancement Tasks

Networking & Coordination Tasks

Project Coordination Tasks

Outputs

Reports for European Commission

Curation Costs Exchange

Collaboration to Clarify the Costs of Curation

Jisc

Reports for General Dissemination

Curation Costs Exchange

Collaboration to Clarify the Costs of Curation
Collaboration to Clarify the Costs of Curation

TIMELINE

- May 2011: Preparation
- Nov 2011: Project Kickoff
- Apr 2012: Archiving
- Feb 2013: Summer 2013
- Summer 2013: Contact Stakeholders
- Summer 2014: Emerging Resources
- Jan 2015: Project Close
- Oct 2014: 4C Conference

Who we are & what we do

- Participation, debate, emerging findings
- Dissemination, legacy, recommendations

- Economics of Digital Curation Roadmap
- Curation Costs Exchange

iPRES 2013

4C Conference
Community Resources

There is a sizeable canon of research into cost modelling for digital curation. This research has tended to emphasize the cost and complexity of digital curation and preservation, but the research is in many ways preliminary and there has been little uptake of the tools and methods that have been developed—for example, tools to manage and estimate costs have not been integrated into other digital curation processes or tools. The question is why? That’s where the 4C project comes in.

The 4C project’s mission is to provide useful, useable resources which support the process of cost management in digital curation. Major outputs from the project will include:
The definition of digital curation ...

The Digital Curation Centre (DCC) definition

Digital curation involves maintaining, preserving and adding value to digital assets throughout their lifecycle.

The 4C Project extended definition

Digital curation involves pre-ingest (appraisal, selection, preparation, rights), ingest, data management, (archival) storage, preservation planning, access, common services, repository administration and general management.
Open Archival Information System (OAIS) Reference Model

ISO 14721:2003
Consultative Committee on Space Data Systems
Why should we concern ourselves about the cost of curation? (What are the stakeholders saying ...)

- Understanding costs can support strategic planning.
- Understanding costs can support tactical decision-making.
- Understanding costs can provide evidence of cost-effectiveness and value.
- Understanding the cost of preservation may mean we can offer realistic and cost effective curation services to others.
- Clarifying and publishing the cost of digital curation can be used to enhance our organisation’s credibility. But this must be done along with the context of how the costs were calculated.
- Understanding economic drivers can help to strategically align an organisation.
Digital Preservation Cost Modelling: Where did it all go wrong?

By paul on 29 June 2012 - 2:08pm

I recently spoke at a workshop on digital preservation costing organised by the lovely people at Knowledge Exchange and Nordbib. After briefly covering some of the work I was previously involved in as part of the LIFE Projects, I talked about why I think that

From the outset, it was *not* our aim to design the definitive cost model that rendered all existing cost models obsolete.

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Acronym</th>
<th>Owner</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Test bed Cost Model for Digital Preservation</td>
<td>T-CMDP</td>
<td>National Archives of the Netherlands</td>
</tr>
<tr>
<td>2</td>
<td>NASA Cost Estimation Tool</td>
<td>NASA-CET</td>
<td>National Aeronautics &amp; Space Administration</td>
</tr>
<tr>
<td>3</td>
<td>LIFE³ Costing Model</td>
<td>LIFE3</td>
<td>University College London and The British Library</td>
</tr>
<tr>
<td>4</td>
<td>Keeping Research DataSafe</td>
<td>KRDS</td>
<td>Charles Beagrie Limited</td>
</tr>
<tr>
<td>5</td>
<td>Cost Model for Digital Archiving</td>
<td>CMDA</td>
<td>Data Archiving and Networked Services (DANS)</td>
</tr>
<tr>
<td>6</td>
<td>Cost Model for Digital Preservation</td>
<td>CMDP</td>
<td>Danish National Archives and The Royal Library, DK</td>
</tr>
<tr>
<td>7</td>
<td>DP4lib Cost Model</td>
<td>DP4lib</td>
<td>German National Library</td>
</tr>
<tr>
<td>8</td>
<td>PrestoPRIME Cost Model for Digital Storage</td>
<td>PP-CMDS</td>
<td>The PrestoPRIME project</td>
</tr>
<tr>
<td>9</td>
<td>Total Cost of Preservation</td>
<td>CDL-TCP</td>
<td>California Digital Library</td>
</tr>
<tr>
<td>10</td>
<td>Economic Model of Long-Term Storage</td>
<td>EMLTS</td>
<td>Rosenthal, D.</td>
</tr>
</tbody>
</table>

Only 15% of people in the 4C stakeholder consultation indicated that they had tried to use a cost model.
We need to think carefully about what exactly is the problem that we are trying to solve ...

If they want to, organisations can work out how much it costs them to manage their digital assets
### 4C Data Gathering Exercise
**Organisation A**

#### Curation Categories
- Pre-Ingest
- Ingest
- Preservation Planning
- Data Management
- Archival Storage
- Access
- Administration

#### Accounting Principles

<table>
<thead>
<tr>
<th>Labour Direct</th>
<th>Labour Indirect</th>
<th>Capital Direct</th>
<th>Capital Indirect</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.2</td>
<td>0.7</td>
<td>0.6</td>
<td>1.3</td>
</tr>
<tr>
<td>0</td>
<td>0.2</td>
<td>0.7</td>
<td>0.6</td>
<td>1.3</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
</tbody>
</table>

- **Time Period:** 2012
- **Total Cost:** €252,000
Collaboration to Clarify the Costs of Curation

4C Data Gathering Exercise
Organisation B

Curation Categories
- Ingest
- Curation
- Access

Cost Categories
- Hardware
- Software
- Employment
- Accommodation
- External Services
- Transfer

Accounting Principles
- Direct
- Costs of Service
- Absorbed Indirect Costs of Service
- Unabsorbed Indirect Costs of Service

Time Period
2012

Total Cost
€645,683.26
Collaboration to Clarify the Costs of Curation

4C Data Gathering Exercise
Organisation C

<table>
<thead>
<tr>
<th>Curation Categories</th>
<th>Accounting Principles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ingest</td>
<td>Labour Costs</td>
</tr>
<tr>
<td>Data Management</td>
<td>Capital Costs</td>
</tr>
<tr>
<td>Archival Storage</td>
<td>Offset By Revenue</td>
</tr>
<tr>
<td>Preservation Planning</td>
<td></td>
</tr>
<tr>
<td>Access</td>
<td></td>
</tr>
<tr>
<td>Administration</td>
<td></td>
</tr>
<tr>
<td>Common Services</td>
<td></td>
</tr>
</tbody>
</table>

Size of Collection
393 TB

Time Period
2012

Total Cost
€15,800,000

<table>
<thead>
<tr>
<th>Category</th>
<th>Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Cost</td>
<td>€15,800,000</td>
</tr>
<tr>
<td>Ingest</td>
<td></td>
</tr>
<tr>
<td>Data Management</td>
<td></td>
</tr>
<tr>
<td>Archival Storage</td>
<td></td>
</tr>
<tr>
<td>Preservation Planning</td>
<td></td>
</tr>
<tr>
<td>Access</td>
<td></td>
</tr>
<tr>
<td>Administration</td>
<td></td>
</tr>
<tr>
<td>Common Services</td>
<td></td>
</tr>
</tbody>
</table>
### Collaboration to Clarify the Costs of Curation

#### 4C Data Gathering Exercise
Organisation D

<table>
<thead>
<tr>
<th>Product/Service</th>
<th>Cost Categories</th>
<th>Time Period</th>
<th>Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hardware</td>
<td></td>
<td>€349,665</td>
</tr>
</tbody>
</table>

Curation Categories

- Ingest
- Archival Storage
- Metadata Management
- Access
- Administration

Cost Categories

- Requirements
- Customisation
- Integration
- Installation
- Training
- Support

Total Cost: €349,665
Collaboration to Clarify the Costs of Curation

4C Data Gathering Exercise
Organisation E

Curation Categories
Digital Archiving

Cost Categories
- Hardware
- Software Maintenance
- Software Development
- Staff/Other

Time Period
- 2007: Total Cost 123,000
- 2012: Total Cost 348,500
Collaboration to Clarify the Costs of Curation

4C Data Gathering Exercise
Organisation F

Curation Categories
Long term Digital Preservation

Size of Collection
2 TB

Assumption
€8k per TB per year for Storage Costs

Accounting Principles
Development & Improvement Operation

Staff
Development, Technical Support
Training, Communications, Public Relations

Expenses
Software Design
Software Licenses
Support
External Development
Hardware Purchase
Hardware Operating costs
Graphic Design

Time Period
2007-2012

Total Cost
€205,000
So ... What exactly are the problems we need to tackle?

- The random numbers problem - How can we meaningfully compare the numbers that we end up with? [cost data]
- Activity based costing versus financial accounting methods
- Describing what the organisation does [cost metadata]
- Describing the amount and type of data that is being looked after [cost metadata]
- Sensitivity around data – Many organisations are not particularly happy to broadcast what it costs them to manage their data. How can we effectively anonymise the sharing of data?
- Complexity - The detail builds up very quickly across different organisations and it doesn’t map together easily
- And we somehow have to make sure that the benefits are presented alongside the costs
So what is needed?

A mechanism for sharing and comparing costs that is a trusted, neutral and sustainable platform

Well! ... What a stroke of luck. The 4C Project just happens to be building one of those
All about the cost of curation.
Understand what you and others are or should be spending in digital curation.

Get started
Introduction to the Curation Cost Exchange, what it does and why.

Compare costs
Add your cost data and see how you compare with others.

Plan investments
Understand how to calculate a budget for future investments by using cost models.

Read more
Browse books, papers and articles to help you get started in curation costing.

Discuss and share
Share your experiences and read about challenges in other organisations like yours.

Find services
See a list of suppliers of digital curation services, what they offer and how much it costs.
We are making a series of (well founded) assumptions about people’s perspectives and people’s behaviour with the Curation Costs Exchange.

One of the things we are charged with doing in the 4C project is to make it easier for people to do cost modelling in the future.

And it’s worth noting:

a) When people approach the problem of trying to work out how much digital curation costs, they may start from different points
   - Which tool should I use?
   - What is a digital curation activity?
   - How do I measure the activity and where does it stop and start?
   - Which calculating method should I use

b) They will probably assert that their organisational context is unorthodox
So what is needed?

Some form of authoritative and comprehensive guidance that is *not* just another cost model *and* is contextually very elastic

Well! ... What a stroke of luck. The 4C Project just happens to be developing something
The 4C Cost Concept Model (CCM) and Gateway Specification
Cost modelling is the primary focus of the CCM and the Gateway specification but it is not an activity that should be done in isolation from other concepts.

An Economic Model for Digital Curation

A representation that describes how economic processes around digital curation work; including the flow of resources (costs and revenues) within the economic lifecycle of digital information assets, and stakeholders interaction with this lifecycle (from the demand, supply and management side).
ECONOMIC MODEL
Maintaining stakeholder incentives and the flow of resources to sustain assets

COSTS MODEL
Activity and Process
Digital Assets
Time and Effort

BENEFITS MODEL
Value & Benefits
Time and Effort
Curation as Black Box

Maintaining stakeholder incentives and the flow of resources to sustain assets.
The 4C Cost Concept Model (CCM) and Gateway Specification
Core concepts are defined simply and in a methodologically-neutral way.

Future cost model theorists and cost model tool developers are urged to use these definitions in their work or to contribute to advancing and improving the definitions.
**Resource**
Direct and indirect cost associated with a particular type of resource, capital or labour.
**Direct cost**
Costs associated with resources used for performing digital curation activities (for example costs of acquisition of storage media, costs of adding metadata), where the amount of resources spent can be directly measured. Also known as variable costs.

**Indirect cost**
Costs incurred by the usage of shared resources, such as general management and administration or common facilities and systems, where it has not been possible to distribute the cost on specific activities. Also known as residual cost or overhead.

**Capital cost**
Cost incurred once, by acquisition (building space, equipment, materials) or by investments. Also known as investment cost or one-time cost.

**Labour cost**
Cost of wages paid to workers.
**Accounting Principles**
Accounting principles regulate the calculations of resources required to complete activities. These principles follow national and international standards. Organisations’ accounting practices will apply these principles but the structure of accounting may not align closely with the needs of cost methodologies.

**Financial Adjustments**
Financial adjustments include inflation (or deflation), depreciation, and interest (discount rates).

**Depreciation/Amortization**
- Depreciation - a mechanism for distributing capital costs over the estimated useful lifetime of a *tangible* asset to indicate how much of the asset’s value has been used.
- Amortization – a mechanism for distributing capital costs over the estimated useful lifetime of an *intangible* asset to indicate how much of the asset's value has been used.
Activity and Services
A measurable amount of work performed by systems and/or people to produce a result
Asset properties
Simple and complex objects (significant properties)

Have we retained a sufficient ‘essence’ of the original?
The Stakeholder Context

“They will probably assert that their organisational context is unorthodox”
The vast majority of institutional decisions will be taken based on calculations which incorporate costs and benefits of a non-financial nature.

These non-financial variables in the decision making process can be characterised as ‘indirect economic determinants.’

They can be regarded as conceptual controls to support discussions at a higher stakeholder or organisational management level.
Exercise - Indirect Economic Determinants
To what extent would your organisation regard the following 15 outcomes as an investment priority?

<table>
<thead>
<tr>
<th>IED</th>
<th>High</th>
<th>Medium</th>
<th>Low</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authenticity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benefit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Efficiency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impact</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Innovation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interoperability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reputation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sustainability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transparency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trustworthiness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Collaboration to Clarify the Costs of Curation

Indirect Economic Determinants

- Business Drivers
- Organisational Objectives
- Management Tools

They are controls that can be used to adjust outcomes within the organisation and they correlate with raised or lowered costs depending on whether you wish to turn them up or down.

- Trustworthiness
- Skills
- Efficiency
- Sustainability
Difficult to Quantify

A desire for a good **reputation, flexibility, and innovation** might be common senior management, policy or funder level goals but may require definition and interpretation before they can be integrated into curation costing calculations.

Measurable with some additional context

**Authenticity** with use of checksums and provenance data. **Impact** might be citations or downloads. **Quality** can be understood at various levels (inc. integrity of retained transformational properties).

Subject to standardised audit and certification

- The ISO9000 series for **quality**
- The ISO31000 for **risk** governance
- ISO27000 for Information Security addresses **authenticity, confidentiality** and **risk**
A conceptual cost and benefit model for digital curation
Part of understanding the organisational context, is to understand all of the stakeholders who have a part to play in the curation of digital assets.

The stakeholder ecosystem

The curation service is an important concept as it denotes an aspiration towards maturity

A curation service will sustain a ‘series of actions’ over time and will enable value to keep on being delivered to relevant stakeholders for as long as the digital assets are capable of realising value.

The issue is ultimately one of sustainability planning ...
Introduction to the Economic Sustainability Reference Model (ESRM)

The ESRM maps out the key elements of the problem space planners face when designing a sustainability strategy for their digital curation activities.

It focuses on the general concept of a sustainability strategy, breaks it down into its key components, and draws planners’ attention to the properties of those components most relevant for economic sustainability.
Collaboration to Clarify the Costs of Curation

Blue Ribbon Task Force on Sustainable Digital Preservation and Access

OCLC  ↓  Jisc

An Economic Sustainability Reference Model

4C Sustainability Self-Assessment Tool

PUBLIC DOMAIN  ↓  ISO

?
What is the purpose of a reference model ...?
Some questions to consider as we go through the session ...

- Does it serve a useful purpose?
- Who would use it?
- Why would they use it?
- When would they use it?
- How would they use it?
- What’s missing?
- How could it be more effective?
Some points to bear in mind as we assess the ESRM + Tool ... 

They are works in progress ... 

They need quite a lot more work done on (both) of them to streamline them and make them efficient and usable. 

Both resources are an attempt to generalise and/or simplify.

It’s tempting (and possible) to jump straight away to the Self-Assessment Tool and try to just fill it in with no reference to the ESRM text. But the concepts in the tool are taken from the text and the text gives more context and meaning to the questions in the tool.
Q6 – Has the stakeholder ecosystem been surveyed or mapped?

For instance ...

Supply-side

Creator

Rights Holder

Managing Agency

Resource Providers

Digital Assets

Demand-side

Current Beneficiaries

Future Beneficiaries

Representatives of the Long-term Public Interest

Lifecycle Management

Select

Creates

Grants right to preserve to Supply resources to

Use

Will use

Act for

Vests IPR in

Creates Use Will use
The ESRM proposes that a sustainability strategy requires consideration of four categories of issues:

• The Economic Lifecycle

• Sustainability Conditions

• Key Entities

• Economic Uncertainties
The activity of digital curation is assumed to be the central active component and the engine that will ensure the sustainability of digital assets.
Investment into curation will in turn facilitate use (or the potential for use)
And use (or the potential for use) will realise value, thereby delivering a return on the investment.
This could play out in a linear fashion with assets being created, curated and then deleted according to a retention schedule. But in the context of sustainability, it is more likely to be a cyclical process.
There will be a gap in the cycle where a technical or business issue introduces a threat to the continued viability of the assets. This becomes a decision point ... Do we allocate more resources to tackling the problem?
Sustainability Conditions

Five Sustainability Conditions are set out to maximise the prospects for sustaining assets
The assets must be understood (or perceived) to have tangible or intangible value
Relevant stakeholders must be sufficiently motivated to support curation
Where resources are scarce then discretion must be used to prioritise curation of the most valuable assets.
The organisation should have an appropriate mandate; a supportive governance structure; and be optimally configured to sustain the assets.
There must be a sufficient flow of ongoing resources (including financial and human capital) to achieve long-term goals.
Key Entities

Three Key Entities are set out which are found in all digital curation contexts. Sustainability requires the nature of these entities to be understood.
Key Entities

Three Key Entities are set out which are found in all digital curation contexts. Sustainability requires the nature of these entities to be understood.

**ASSETS**

Every type of digital asset exhibits various attributes or properties that to a greater or lesser extent may affect the how they are curated.

**STAKEHOLDERS**

The stakeholder ecosystem for digital assets can be complex and the supply side and demand side should be understood in relation to who is undertaking the curation for the benefit of whom.

**PROCESSES**

The processes involved must be capable of (and optimised for) efficiently enhancing the value of the assets.
The inclusion of Economic Uncertainties is an acknowledgement that even the best sustainability strategy cannot accurately predict the future and that some expectation or mitigation of uncertainty (both threats and opportunities) should be built into the strategy where possible.
The Costs of Curation – are we on the right track?

Neil Grindley
n.grindley@jisc.ac.uk
Head of Resource Discovery

Coordinator,
The 4C Project

4th July 2014, Split, Croatia

Thanks for coming!